EVIDENCE FOR BENEFIT?

REVIEWING EMPIRICAL RESEARCH ON THE USE OF DIGITAL TOOLS IN MATHEMATICS EDUCATION

Paul Drijvers

Freudenthal Institute, Utrecht University, Utrecht, The Netherlands

A recent OECD study has raised the question of which evidence we have for the benefit of using digital technology in mathematics education. To investigate this, we focus on experimental and quantitative studies, and revisit the main recent review studies in this domain. The results show significant positive effects of the use of digital technology on student achievement but only small effect sizes. As a conclusion, a plea is made for replication studies and for studies that identify decisive factors through the combination of a methodologically rigorous design and a theoretical foundation in domain-specific theories from mathematics didactics.

INTRODUCTION

"Computers 'do not improve' pupil results, says OECD" was the header of a September 2015 BBC news item[1]. A Dutch news site[2] provided even stronger a claim: "Poorer school performance through increased computer use." Both items were based on the publication of a report on student achievement and the use of computers (OECD, 2015). Indeed, the results of this study show negative correlations between mathematics performance and computer use in mathematics lessons and lead to conclude that there is little evidence for a positive effect on student achievement:

> Despite considerable investments in computers, internet connections and software for educational use, there is little solid evidence that greater computer use among students leads to better scores in mathematics and reading. (OECD, 2015, p. 145)

However, correlations do not imply causality. For the case of mathematics education, NCTM claims that we cannot neglect digital tools: "Technology is an essential tool for learning mathematics in the 21st century" (NCTM, 2008, p. 1). This quote reflects the optimism concerning the potential of digital technology for mathematics teaching and learning, including a possibly changing focus in mathematics curricula towards conceptual understanding and higher order thinking skills. This optimism is underpinned by research findings, such as the ones reported by Ronau et al. (2014):

> Over the last four decades, research has led to consistent findings that digital technologies such as calculators and computer software improve student understanding and do no harm to student computational skills. (Ronau et al., 2014, p. 974)

These different claims and opinions challenge the research community to be clear about what is known about the effects of using digital technology in mathematics education on student performance. The important work done by OECD, with its non-experimental character, is not the

[1] 15 september 2015, http://www.bbc.com/news/business-34174796
[2] 15 september 2015, http://nos.nl/artikel/2057772-slechtere-schoolprestaties-door-meer-computergebruik.html

final word. Therefore, this contribution to TSG42, and to the theme "evidence for benefit" in particular, addresses this topic by investigating the results of empirical, quantitative studies on the use of digital technology in mathematics education from a reflective stance.

First, we will revisit and synthesize the results of some important studies in the field, with a particular focus on recent review studies. Next, an example of an empirical study that was not successful will be described, as well as some possible causes of the lack of improvement in student performance. In the reflection section, the subtlety of using digital tools in mathematics education will be addressed and some methodological issues are discussed. In the conclusion, I make a plea for methodologically rigorous studies grounded in theories on the learning of mathematics.

REVISTING RESEARCH

A sketch of previous work before 2010

The question of the benefits of integrating digital tools in mathematics education, of course, is not new and has been investigated before. Heid (1997) provides an overview of principles and issues of the integration of digital technology, and sketches the landscape of the different types of tools and their pedagogical potential. Burrill et al. (2002) report on 43 studies on the use of handheld graphing technology and conclude that these devices can be an important factor in helping students develop a better understanding of mathematical concepts; this conclusion, however, is not quantitatively underpinned. Ellington (2003, 2006) also focuses on graphing calculators. Her review of 54 studies shows an improvement of students' operational skills and problem-solving skills when calculators are an integral part of testing and instruction, but with small effect sizes. Lagrange, Artigue, Laborde and Trouche (2003) developed a multi-dimensional framework to review a corpus of 662 research studies on the use of technology in mathematics education and to investigate the evolution of research in the field but did not explicitly address learning outcomes. Kulik (2003) does explicitly address learning outcomes and reports an average effect size of d = 0.38 in 16 studies on the effectiveness of integrated learning systems in mathematics[3]. Two subsequent large-scale experimental studies by Dynarski et al. (2007) and Campuzano, Dynarski, Agodini and Rall (2009), however, conclude that the effects of the use of digital tools in grade 9 algebra courses was not statistically different from zero. Specifically for the use of computer algebra systems, Tokbah (2008) finds significant positive effects with an average of d = 0.38 over 102 effect sizes. Altogether, these studies provide mixed findings on the effect of using digital tools in mathematics education and show different degrees of quantitative evidence.

Three recent review studies

Let us now focus on three more recent review studies that provide information on the effect of using digital technology in mathematics education through reporting effect sizes. The first one is the study by Li and Ma (2010). It reviews 46 studies on using computer technology on mathematics education in K-12 classrooms, reporting in total 85 effect sizes. The researchers found a statistically significant effect with a weighted average effect size of d = 0.28. The reported unweighted average effect size, d = 0.71, seems less appropriate as it doesn't take into account the number of students involved. Additional findings were that higher effect sizes were found in primary education

[3] This means that the average difference between experiment group and control group equals 0.38 of their pooled standard deviation, which can be considered a weak to medium effect.

compared to secondary, and in special education compared to general education. Also, effect sizes were bigger in studies that used a constructivist approach to teaching, and in studies that used non-standardized tests.

The second review study by Rakes, Valentine, McGatha, and Ronau (2010) focuses on algebra in particular. The authors used strict criteria to include studies, which led to 109 effect sizes. The interventions were categorized; here we only report on the categories Technology tools and Technology curricula. The average weighted effect sizes for these two categories were $d = 0.151$ and $d = 0.165$, respectively. Over all categories, the authors concluded that interventions focusing on conceptual understanding provide about twice as high effect sizes as the interventions focusing on procedural understanding. Also, they noted that interventions over a small period of time may have significant effect, and that the grain size differences in interventions (whole-school study versus single-teacher interventions) did not make a significant difference.

The third review study by Cheung and Slavin (2013) set consistent standards for studies to be included. The average of 74 effect sizes in K-12 mathematics studies was $d = 0.16$. The authors' final conclusion refers to a modest difference: "Educational technology is making a modest difference in learning of mathematics. It is a help, but not a breakthrough." (Cheung & Slavin, 2013, p. 102). Some additional findings are worth mentioning. First, in spite of what might be expected due to the development of sophisticated tools, improvement of ICT infrastructure and growing pedagogical experience, the overall effectiveness of educational technology did not improve over time. Second, like Li and Ma (2010), the authors found higher effect sizes in primary than in secondary education. Third, lower effect sizes were found in randomized experiments compared to quasi-experimental studies. Fourth and final, effect sizes in studies with a large number of students were smaller than in small-scale studies.

Study	Number of effect sizes	Average effect size	Global conclusion
Li & Ma 2010	85	0.28 (weighted)	Statistically significant positive effects.
Rakes et al. 2010	109	0.151 – 0.165	Positive, statistically significant results.
Cheung & Slavin 2013	74	0.16	A positive, though modest effect

Table 1: Effect sizes reported in three review studies

Table 1 summarizes the findings of the three review studies. The overall image is that the use of digital technology in mathematics education can have a significant positive effect, but with a small effect size. As such, these studies do not provide an overwhelming evidence for the effectiveness of the use of digital tools in mathematics education. Also, as will be shown in the next section, studies that report a lack of benefit continue to be published as well.

AN EXAMPLE: THE CASE OF APPLETS FOR ALGEBRA

To illustrate that providing evidence for the benefits of using digital tools is not straightforward, I refer to a study on the use of two online algebra modules in 8th grade (Drijvers, Doorman, Kirschner, Hoogveld, & Boon, 2014). The modules were designed in the Digital Math Environment, which proved to be successful in improving student achievement in algebra in grade 12 (Bokhove & Drijvers, 2012). Also, teachers had reported success while implementing the online materials in lower grades. Figure 1 shows a task from one of the modules.

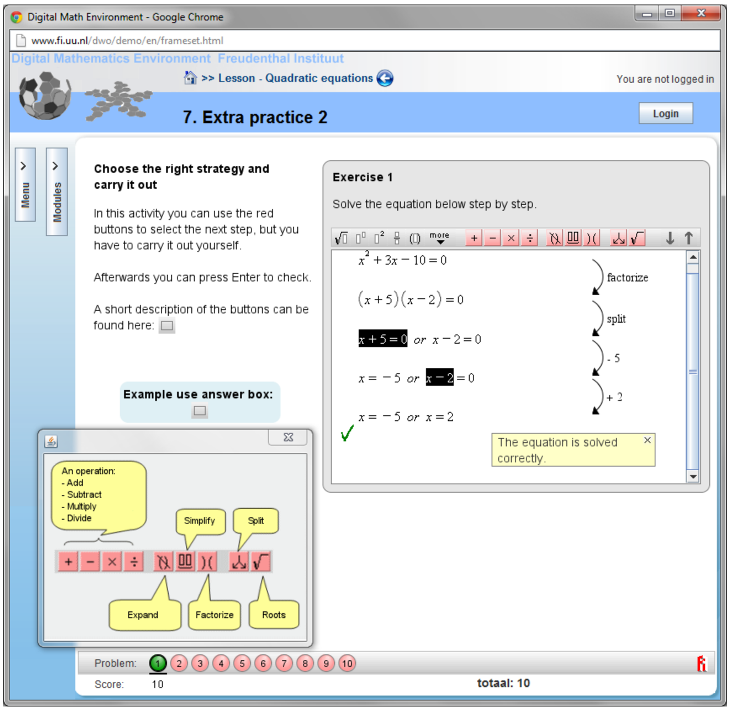

Figure 1: Algebra task in the Digital Math Environment

The study had an experimental design, in which each of the involved teachers taught to two classes in parallel, each randomly assigned to the experimental condition of using the online modules, or to the control condition of regular teaching. Figure 2 shows the results of the pretest, the intermediate test (Post_Linear), the posttest (Post_Quadractic) and the two retention tests, all administered with paper-and-pen. In spite of the earlier positive experiences with these types of modules, the results show that the experimental condition did not lead to students outperforming their peers in the control condition. The experimental group did not catch up the small initial (and coincidental) lag; indeed, this gap got significantly larger in the final retention test!

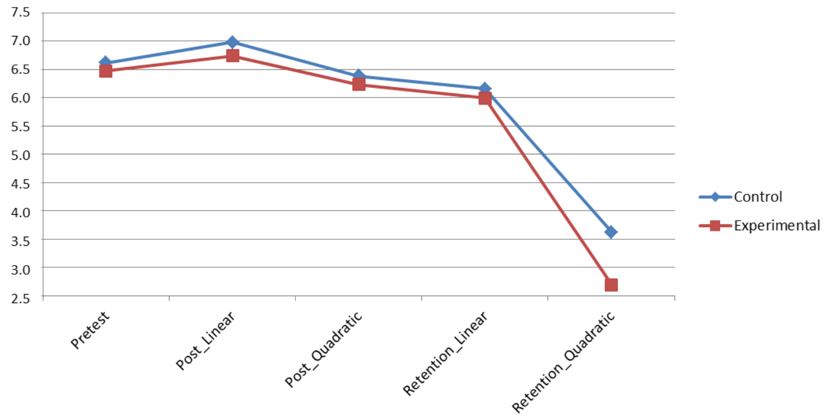

Figure 2: Average grades for control and experimental group (N = 842)

As possible explanations for these findings, the authors mention a spill-over effect. All participating teachers taught one control and one experimental class, and they may have picked up pedagogical ideas from the online intervention and used these in the control classes as well. A second possible explanation is that the work on the online tasks was not an adequate preparation for more complex tasks. Third, the feedback provided in the digital environment might have lacked quality, and, finally, the integration of paper-and-pen skills and digital practice might not have been optimal, so that transfer to the “traditional media” was hindered. In short, in spite of a careful experimental design and an environment that had proven to be useful in other settings, the researchers did not find a positive effect.

REFLECTING ON THE STATE OF THE ART

If we reflect on the above finding, we first notice that the results from experimental and quantitative studies are more positive than the correlational findings from the OECD study. However, effect sizes are limited, all the more if we realize that any educational innovative intervention usually has a positive effect anyway (Higgins et al., 2012). Like for *p*-values[4], the interpretation of effect sizes should be done with care. So the above review does not completely solve the issue and the picture is still diffuse. How to explain this? My feeling is that two factors play an important role here: the too general claims, that ignore the subtlety of using digital tools for learning, and the methodological weaknesses that some studies suffer from. Let me address the two in more detail.

The too general claims that ignore the subtlety of using digital tools for learning

Wouldn’t we all agree that research findings such as “The use of paper-and-pen has a positive effect on student achievement” would be too general? Then why do we try to find evidence for similar claims for the case of ICT? It makes sense to assume that digital technology is not a panacea, and that its effectiveness will depend upon particular implementations and situations. The following two quotes show that the effect of ICT in mathematics education is a subtle matter and will depend to an important extent on the specific technological application, the educational setting and the orchestration by the teacher. It is the “how” that counts!

[4] For a current debate on p-values see http://www.statslife.org.uk/news/2116-academic-journal-bans-p-value-significance-test

> Studies linking the provision and use of technology with attainment tend to find consistent but small positive associations with educational outcomes. However a causal link cannot be inferred from this kind of research. It seems probable that more effective schools and teachers are more likely to use digital technologies more effectively than other schools. We need to know more about where and how it is used to greatest effect, then investigate to see if this information can be used to help improve learning in other contexts. We do not know if it is the use of technology that is making the difference. (Higgins, Xiao & Katsipataki, 2012, p. 3)

> There have been several reviews of the benefits of ICT to student learning in mathematics that suggest positive effects from the use of digital technology. […] However, the type and extent of the gains are a function of how the technology is used in the teaching of mathematics. (Drijvers, Monaghan, Thomas, & Trouche, 2015, p. 15).

If we agree that the learning of mathematics is a complex domain and that we need to know more about the factors that determine the contribution of digital tools to it, it is important that research is grounded in theoretical knowledge from domain-specific mathematics pedagogy and from man-machine interaction. Theories on reification (Sfard, 1991) or on instrumental genesis, for example, may offer such a theoretical basis (e.g., see Drijvers, Kieran & Mariotti, 2010). Educational research on the use of digital tools for mathematics education that is not based upon domain-specific didactical knowledge may miss opportunities to discover decisive factors.

As an aside, we should note that didactical knowledge and practice may also change under the influence of digital technology. In fact, this is what OECD (2015, p. 3-4) mentions as a possible explanation for their surprising findings: "… we have not become good enough at the kind of pedagogies that make the most of technology. […] . Technology can amplify great teaching but great technology cannot replace poor teaching."

Methodological weaknesses

An overview of the field shows some remarkable methodological characteristics. First, replication studies are hardly ever carried out. Why is this the case? If we would have replication studies, would we do better than such studies in the field of cognitive and social psychology[5]?

Second, it is interesting to notice that smaller studies tend to report bigger effect sizes than bigger ones (Cheung & Slavin, 2013) and that the reported effect sizes do not seem to increase over time. This suggests that the effective small-scale studies may not be so easy to scale up. As far as the trend over time is concerned, the criteria for publication and for inclusion in review studies seem to be increasing. Maybe these higher methodological standards filter out some studies that report high effect sizes?

From a methodological point, more rigor in research methods may improve the quality of our results. From this perspective, I agree with Ronau and colleagues, who in a recent study on the quality of 480 mathematics education technology dissertations make plea for a higher quality in both research reports and reviews:

> The mathematics education technology research community must in turn begin to demand greater quality in its published studies, through both how researchers write about their own studies and how they review the works of others. (Ronau et al., 2014, p. 1002)

[5] See, for example, http://www.theguardian.com/science/2015/aug/27/study-delivers-bleak-verdict-on-validity-of-psychology-experiment-results

CONCLUSION

In the introduction, we raised the question of what is known about the effects of using digital technology in mathematics education on student performance. The literature review reveals mixed results. The OECD correlational study shows little evidence for benefit. Experimental studies, and their review studies in particular, report significant and positive effects, but with small average effect sizes in the order of $d = 0.2$. Compared to effect sizes reported for other types of innovative interventions, this is not much. Also, insight into factors that are decisive for the (lack of) positive benefit of the use of digital tools is limited.

Of course, the above approach has some important limitations. First, review studies are based on studies that are older themselves, and one might wonder if the picture has changed over, say, the last five years. The fact that effect sizes so far have not been increasing does not favour this argument. Second, we focus on experimental, quantitative studies and neglect qualitative studies and studies that follow a design research paradigm, whereas such studies can contribute to the body of knowledge and in many cases take an in-depth view on student learning and are firmly grounded in theories from the field of mathematics didactics[6]. A third limitation of the type of review studies revisited, is that these studies do not differentiate between educational level, type of technology used, and other educational factors that may be decisive. Rather, they provide an overview without nuances, which may cause us to miss important insights in the phenomenon.

In spite of these limitations, the conclusion is that evidence for the benefit of using technology in mathematics education from experimental studies is not very strong. What we need on our research agendas are studies (including replication studies) that focus on the identification of decisive factors that determine the eventual benefits. Such studies should on the one hand be methodologically well designed according to the standards from educational science, and on the other hand be strongly based in sound theoretical foundations from domain-specific mathematics didactics. To combine the best of both worlds is the challenge we are facing.

REFERENCES

Bokhove, C., & Drijvers, P. (2012). Effects of a digital intervention on the development of algebraic expertise. *Computers & Education, 58*(1), 197-208.

Burrill, G., Allison, J., Breaux, G., Kastberg, S., Leatham, K., & Sanchez, W. (2002). *Handheld graphing technology at the secondary level: Research findings and implications for classroom practice*. Dallas, TX: Texas Instruments Corp. http://education.ti.com/research.

Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts – Executive summary* (NCEE 2009-4042). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review, 9*, 88-113.

Drijvers, P., Doorman, M., Kirschner, P., Hoogveld, B., & Boon, P. (2014). The effect of online tasks for algebra on student achievement in grade 8. *Technology, Knowledge and learning, 19*, 1-18.

[6] The findings from qualitative studies are addressed in Heid's contribution to ICME13 TSG42.

Drijvers, P., Kieran, C., & Mariotti, M. A. (2010). Integrating technology into mathematics education: theoretical perspectives. In C. Hoyles, & J.-B. Lagrange (Eds.), *Mathematics education and technology - Rethinking the terrain* (pp. 89-132). New York/Berlin: Springer.

Drijvers, P., Monaghan, J., Thomas, M., & Trouche, L. (2015). *Use of Technology in Secondary Mathematics. Final Report for the International Baccalaureate*. http://www.ibo.org/globalassets/publications/ib-research/technologyindpmathematicsfinalreport.pdf.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., & Campuzano, L. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort. Report to Congress. Publication NCEE 2007-4005*. Washington, DC: U.S. Department of Education.

Ellington, A. J. (2003). A meta-analysis of the effects of calculators on students' achievement and attitude levels in precollege mathematics classes. *Journal for Research in Mathematics Education, 34*(5), 433-463.

Ellington, A. J. (2006) The Effects of non-CAS graphing calculators on student achievement and attitude levels in mathematics: A meta-analysis. *School Science and Mathematics, 106*(1), 16-26.

Heid, M. K. (1997). The Technological Revolution and the Reform of School Mathematics. *American Journal of Education*, *106*(1), 5-61.

Higgins, S., Xiao, Z., & Katsipataki, M. (2012). *The Impact of Digital Technology on Learning: A Summary for the Education Endowment Foundation*. Report. Durham: Durham University.

Kulik, J. A. (2003). Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say. Arlington, VA: SRI International. http://www.sri.com/policy/csted/reports/sandt/it/Kulik_ITinK-12_Main_Report.pdf.

Lagrange, J.-B., Artigue, M., Laborde, C., & Trouche, L. (2003). Technology and Mathematics Education: a Multidimensional Study of the Evolution of Research and Innovation. In A. J. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Second International Handbook of Mathematics Education* (pp. 239-271). Dordrecht, Netherlands: Kluwer Academic Publishers.

Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review, 22*, 215-243.

National Council of Teachers of Mathematics (2008). *Technology in Teaching and Learning Mathematics. A position of the National Council of Teachers of Mathematics*. http://www.nctm.org/Standards-and-Positions/Position-Statements/Technology-in-Teaching-and-Learning-Mathematics/ .

Rakes, C. R., Valentine, J. C., McGatha, M. B., & Ronau, R. N. (2010). Methods of instructional improvement in Algebra: A systematic review and meta-analysis. *Review of Educational Research, 80*(3), 372-400.

Ronau, R. N., Rakes, C. R., Bush, S. B., Driskell, S. O., Niess, M. L., & Pugalee, D. K. (2014). A Survey of Mathematics Education Technology Dissertation Scope and Quality: 1968–2009. *American Educational Research Journal*, *51*(5), 974–1006.

OECD (2015). *Students, Computers and Learning. Making the Connection*. Paris: OECD Publishing. DOI:10.1787/9789264239555-en. http://www.oecd.org/edu/students-computers-and-learning-9789264239555-en.htm.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics, 22*(1), 1–36.

Tokpah, C. L. (2008). *The Effects of Computer Algebra Systems on Students' Achievement in Mathematics*. Doctoral dissertation. Kent, OH: Kent State University. https://etd.ohiolink.edu/!etd.send_file?accession=kent1215528446&disposition=inline.